AMENDMENTS TO THE CLAIMS:

This listing of claims will replace all prior versions and listings of claims in the application:

Listing of Claims:

1. **(Currently amended)** A method of clustering documents [[(]] or patterns [[)]] each having one or plural document [[(]] or pattern [[)]] segments in an input document [[(]] or pattern [[)]] set, said method comprising: ~~based on a relation among them, comprising,~~

(a) obtaining a document [[(]] or pattern [[)]] frequency matrix for the set of input documents [[(]] or patterns [[)]], based on occurrence frequencies of terms appearing in each document [[(]] or pattern [[)]];

(b) selecting a seed document [[(]] or pattern [[)]] from remaining documents [[(]] or patterns [[)]] that are not included in any cluster existing at that moment and constructing a current cluster of the initial state using the seed document [[(]] or pattern [[)]];

(c) obtaining the document [[(]] or pattern [[)]] commonality to the current cluster for each document [[(]] or pattern [[)]] in the input document [[(]] or pattern [[)]] set by using information based on the document [[(]] or pattern [[)]] frequency matrix for the input document [[(]] or pattern [[)]] set, information based on the document [[(]] or pattern [[)]] frequency matrix for documents [[(]] or patterns [[)]] in the current cluster and information based on the common co-occurrence matrix of the current cluster, and making documents [[(]] or patterns [[)]] having the document commonality higher than a threshold belong temporarily to the current cluster;

(d) repeating step (c) until the number of documents [[(]] or patterns [[)]] temporarily belonging to the current cluster becomes the same as that in the previous repetition;

(e) repeating steps (b) through (d) until a given convergence condition is satisfied; and

(f) deciding, on the basis of the document [[(]] or pattern [[)]] commonality of each document [[(]] or pattern [[)]] to each cluster, a cluster to which each document [[(]] or pattern [[)]] belongs and outputting said cluster.

2. **(Currently amended)** A clustering method according to claim 1, wherein step (a) further includes: [[,]]

(a-1) generating a document [[(]] or pattern [[)]] segment vector for each of said document [[(]] or pattern [[)]] segments based on occurrence frequencies of terms appearing in each document [[(]] or pattern [[)]] segment;

(a-2) obtaining a co-occurrence matrix for each document [[(]] or pattern [[)]] in the input document [[(]] or pattern [[)]] set from the document [[(]] or pattern [[)]] segment vectors; and

(a-3) obtaining a document [[(]] or pattern [[)]] frequency matrix from the co-occurrence matrix for each document.

3. **(Currently amended)** A clustering method according to claim 1, wherein step (b) further includes: [[,]]

(b-1) constructing a common co-occurrence matrix of remaining documents [[(]] or patterns [[)]] that are not included in any cluster existing at that moment; and

(b-2) obtaining a document commonality to the set of the remaining document [[(]] or pattern [[)]] set for each document [[(]] or pattern [[)]] in the remaining document [[(]] or pattern [[)]] set by using the common co-occurrence matrix of the remaining documents [[(]] or patterns [[)]], and extracting the document [[(]] or pattern [[)]] having the highest document [[(]] or pattern [[)]] commonality, and constructing a current cluster of the initial state by making a document [[(]] or pattern [[)]] set including the seed document [[(]] or pattern [[)]] and the neighbor documents [[(]] or patterns [[)]] similar to the seed document [[(]] or pattern [[)]].

4. (**Currently amended**) A clustering method according to claim 1, wherein step (c) further includes: [[,]]

(c-1) constructing a common co-occurrence matrix of the current cluster and a document [[(]] or pattern [[)]] frequency matrix of the current cluster;

(c-2) obtaining the distinctiveness of each term and each term pair to the current cluster by comparing the document [[(]] or pattern [[)]] frequency matrix of the input document [[(]] or pattern [[)]] set and the document [[(]] or pattern [[)]] frequency matrix of the current cluster; and

(c-3) obtaining document [[(]] or pattern [[)]] commonalities to the current cluster for each document [[(]] or pattern [[)]] in the input document [[(]] or pattern [[)]] set by using the common co-occurrence matrix of the current cluster and weights of each term and term pair obtained from their distinctiveness, and making a document [[(]] or pattern [[)]] having the document [[(]] or pattern [[)]] commonality higher than a threshold belong temporarily to the current cluster.

5. **(Currently amended)** A clustering method according to claim 1, further including: [[,]]

repeating step (e) until the number of documents [[(]] or patterns [[)]] whose document [[(]] or pattern [[)]] commonalities to any current clusters are less than a threshold becomes 0, or the number is less than a threshold and is equal to that of the previous repetition.

6. **(Currently amended)** A clustering method according to claim 1, wherein step (f) further includes: [[,]]

checking existence of a redundant cluster, and removing, when the redundant cluster exists, the redundant cluster and again deciding the cluster to which each document belongs.

7. **(Currently amended)** A method according to claim 1, wherein the co-occurrence matrix S^r of the document [[(]] or pattern [[)]] D_r is determined in accordance with:

$$S^r = \Sigma_{y=1}^{Y_r} d_{ry} d_{ry}^T \qquad (1)$$

where: M equals the number of sorts of the occurring terms, D_r equals the *rth* document [[(]] or pattern [[)]] in a document [[(]] or pattern [[)]] set D consisting of R documents [[(]] or patterns [[)]], Y_r equals the number of document [[(]] or pattern [[)]] segments in document [[(]] or pattern [[)]] D_r, and $d_{ry} = [[(]] d_{ry1}, .., d_{ryM} [[)]]^T$ equals the *yth* document [[(]] or pattern [[)]] segment vector of document [[(]] or pattern [[)]] D_r, and T represents transposition of a vector.

8. **(Currently amended)** A method according to claim 1, wherein each component of the document [[(]] or pattern [[)]] frequency matrix of a document [[(]] or pattern [[)]] set D is

the number of documents [[(]] or patterns [[)]] in which a corresponding component of the co-occurrence matrix of each document [[(]] or pattern [[)]] in the document [[(]] or pattern [[)]] set D does not take a vale of zero.

9. **(Currently amended)** A method according to claim 1, further comprising:

determining the common co-occurrence matrix of a document [[(]] or pattern [[)]] set D from a matrix T^A on the basis of a matrix T whose mn component is determined by the matrix T^A having an mn component determined by

$T^A_{mn} = T_{mn}$, $U_{mn} > A$,

$T^A_{mn} = 0$ otherwise,

where U_{mn} represents the mn component of the document [[(]] or pattern [[)]] frequency matrix of the document [[(]] or pattern [[)]] set D.

10. **(Currently amended)** A method according to claim 1, further comprising:

determining the common co-occurrence matrix of a document [[(]] or pattern [[)]] set D from a matrix Q^A on the basis of a matrix T whose mn component is determined by

$$T_{mn} = \prod_{\substack{r=1 \\ S^r{}_{mn}>0}}^{R} S^r{}_{mn}$$

the matrix Q^A having an mn component determined by

$Q^A_{mn} = \log[[(]] T^A_{mn} [[)]]$ $T^A_{mn} > 1$,

$Q^A_{mn} = 0$ otherwise.

11. **(Currently amended)** A method according to claim 10_a wherein z_{mm} and z_{mn} are respectively weights for a term [[(]] or object feature [[)]] m and a term [[(]] or object feature [[)]] pair m, n, a document [[(]] or pattern [[)]] commonality of document [[(]] or pattern [[)]] P having a co-occurrence matrix S^P with respect to the document [[(]] or pattern [[)]] set D given by

$$com_l(D,P;Q^A)=\frac{\sum_{m=1}^{M} z_{mm}Q^A{}_{mm}S^P{}_{mm}}{\sqrt{\sum_{m=1}^{M} z_{mm}(Q^A{}_{mm})^2}\sqrt{\sum_{m=1}^{M} z_{mm}(S^P{}_{mm})^2}} \qquad (3)$$

or

$$com_q(D,P;Q^A)=\frac{\sum_{m=1}^{M}\sum_{n=1}^{M} z_{mn}Q^A{}_{mn}S^P{}_{mn}}{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{M} z_{mn}(Q^A{}_{mn})^2}\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{M} z_{mn}(S^P{}_{mn})^2}} \qquad (4).$$

12. **(Currently amended)** A method according to claim 9_a wherein z_{mm} and z_{mn} are respectively weights for a term [[(]] or object feature [[)]] m and a term [[(]] or object feature [[)]] pair m, n, a document [[(]] or pattern [[)]] commonality of document [[(]] or pattern [[)]] P having a co-occurrence matrix S^P with respect to the document [[(]] or pattern [[)]] set D given by

$$com_l(D,P;T^A)=\frac{\sum_{m=1}^{M} z_{mm}T^A{}_{mm}S^P{}_{mm}}{\sqrt{\sum_{m=1}^{M} z_{mm}(T^A{}_{mm})^2}\sqrt{\sum_{m=1}^{M} z_{mm}(S^P{}_{mm})^2}} \qquad (3)$$

or

$$com_q(D,P;T^A)=\frac{\sum_{m=1}^{M}\sum_{n=1}^{M} z_{mn}T^A{}_{mn}S^P{}_{mn}}{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{M} z_{mn}(T^A{}_{mn})^2}\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{M} z_{mn}(S^P{}_{mn})^2}} \qquad (4).$$

13. **(Currently amended)** A method according to claim 1, wherein extraction of the seed document [[(]] or pattern [[)]] of the current cluster and construction of the current cluster of the initial state ~~includes~~ comprise:

(a') obtaining a document [[(]] or pattern [[)]] commonality to the remaining document [[(]] or pattern [[)]] set for each document [[(]] or pattern [[)]] in the remaining document [[(]] or pattern [[)]] set by using the said common co-occurrence matrix of the remaining documents [[(]] or patterns [[)]],

(b') extracting, as candidates of the seed of the current cluster, a specific number of documents [[(]] or patterns [[)]] whose document [[(]] or pattern [[)]] commonalities obtained by step (a') are large;

(c') obtaining similarities of the respective candidates of the seed of the cluster to all documents [[(]] or patterns [[)]] in the input document [[(]] or pattern [[)]] set or in the remaining document [[(]] or pattern [[)]] set, and obtaining documents [[(]] or patterns [[)]] having similarities larger than a threshold as neighbor documents [[(]] or patterns [[)]] of the candidate; and

(d') selecting the candidate whose number of the neighbor documents [[(]] or patterns [[)]] is the largest among the candidates as the seed of the current cluster and making its neighbor documents [[(]] or patterns [[)]] the current cluster of the initial state.

14. **(Currently amended)** A method according to claim 1, further including:

detecting the distinctiveness of each term [[(]] or object feature [[)]] and each term pair with respect to the current cluster and detecting their weights,

the distinctiveness and weight detecting steps including:

(a") obtaining a ratio of each component of a document [[(]] or pattern [[)]] frequency matrix obtained from the input document [[(]] or pattern [[)]] set to a corresponding component of a document [[(]] or pattern [[)]] frequency matrix obtained from the current cluster as a document [[(]] or pattern [[)]] frequency ratio of each term [[(]] or feature [[)]] or each term [[(]] or feature [[)]] pair;

(b") selecting a specific number of terms [[(]] or features [[)]] or term [[(]] or feature [[)]] pairs having the smallest document [[(]] or pattern [[)]] frequency ratios among a specific number of terms [[(]] or features [[)]] or term [[(]] or feature [[)]] pairs having the highest document [[(]] or pattern [[)]] frequencies, and obtaining the average of the document [[(]] or pattern [[)]] frequency ratios of the selected terms [[(]] or features [[)]] or term [[(]] or feature [[)]] pairs as the average document [[(]] or pattern [[)]] frequency ratio;

(c") dividing the average document [[(]] or pattern [[)]] frequency ratio by the document [[(]] or pattern [[)]] frequency ratio of each term [[(]] or feature [[)]] or each term [[(]] or feature [[)]] pair as a measure of the distinctiveness of each term [[(]] or feature [[)]] or each term [[(]] or feature [[)]] pair; and

(d") determining the weight of each term [[(]] or feature [[)]] or each term [[(]] or feature [[)]] pair from a function having the distinctiveness measure as a variable.

15. **(Currently amended)** A method according to claim 1, further including:

eliminating terms [[(]] or features [[)]] and term [[(]] or feature [[)]] pairs having document [[(]] or pattern [[)]] frequencies higher than a threshold.

16. **(Currently amended)** A method according to claim 1, wherein clustering is performed recursively by letting the document **[[(]]** or pattern **[[)]]** set included in a cluster be the input document **[[(]]** or pattern **[[)]]** set.

17. **(Currently amended)** A computer program product containing a computer program which, when executed by a computer, causes the ~~for causing a~~ computer to perform the method of claim 1.

18. **(Currently amended)** A computer program product containing a computer program which, when executed by a computer, causes the ~~for causing a~~ computer to perform the method of claim 2.

19. **(Currently amended)** A computer program product containing a computer program which, when executed by a computer, causes the ~~for causing a~~ computer to perform the method of claim 3.

20. **(Currently amended)** A computer program product containing a computer program which, when executed by a computer, causes the ~~for causing a~~ computer to perform the method of claim 4.

21. **(Currently amended)** A computer program product containing a computer program which, when executed by a computer, causes the ~~for causing a~~ computer to perform the method of claim 5.

22. **(Currently amended)** A computer program product containing a computer program which, when executed by a computer, causes the ~~for causing a~~ computer to perform the method of claim 6.

23. (Original) A computer arranged to perform the method of claim 1.

24. (Original) A computer arranged to perform the method of claim 2.

25. (Original) A computer arranged to perform the method of claim 3.

26. (Original) A computer arranged to perform the method of claim 4.

27. (Original) A computer arranged to perform the method of claim 5.

28. (Original) A computer arranged to perform the method of claim 6.

29. **(Currently amended)** A clustering apparatus for clustering documents [[(]] or patterns [[)]] each having one or plural document [[(]] or pattern [[)]] segments in an input document [[(]] or pattern [[)]] set ~~based on the relation among them~~, the apparatus comprising:

[[(a)]] a first unit [[means]] for obtaining a document [[(]] or pattern [[)]] frequency matrix for the set of input documents [[(]] or patterns [[)]], based on occurrence frequencies of terms appearing in each document [[(]] or pattern [[)]];

[[(b)]] a second unit [[means]] for selecting a seed document [[(]] or pattern [[)]] from remaining documents [[(]] or patterns [[)]] that are not included in any cluster existing at that moment and constructing a current cluster of the initial state using the seed document [[(]] or pattern [[)]];

[[(c)]] a third unit [[means]]

for obtaining the document [[(]] or pattern [[)]] commonality to the current cluster for each document [[(]] or pattern [[)]] in the input document [[(]] or pattern [[)]] set using information based on the document [[(]] or pattern [[)]] frequency matrix for the input document [[(]] or pattern [[)]] set, information based on the document [[(]] or pattern [[)]] frequency matrix for documents [[(]] or patterns [[)]] in the current cluster and information based on the common co-occurrence matrix of the current cluster, and [[means]]

for making documents [[(]] or patterns [[)]] having the document [[(]] or pattern [[)]] commonality higher than a threshold belong temporarily to the current cluster;

[[(d)]] a fourth unit [[means]] for repeating the operations of the third unit ~~means (c)~~ until the number of documents [[(]] or patterns [[)]] temporarily belonging to the current cluster becomes the same as that in the previous repetition;

[[(e)]] a fifth unit [[means]] for repeating the operations of the second through fourth units ~~means (b) through (d)~~ until given convergence conditions are satisfied; and

[[(f)]] a sixth unit [[means]] for deciding, on the basis of the document [[(]] or pattern [[)]] commonality of each document [[(]] or pattern [[)]] to each cluster, a cluster to which each document [[(]] or pattern [[)]] belongs, and for outputting said cluster.